AL/PS-TR-1997-0165

UNITED STATES AIR FORCE
ARMSTRONG LABORATORY

MODIFICATIONS TO THE AIR FORCE CYCLE ERGOMETRY PROTOCOL: IMPACT ON PASS, FAIL, AND INVALID OUTCOMES

Pete Flatten
Dean Richardson
Gerald De Wolfe
Susan Chao
Stefan Constable
Melissa Hite
Ryan O'dowd

OFFICE FOR PREVENTION AND HEALTH
SERVICES ASSESSMENT (OPHSA)
2602 Doolittle Road
Brooks Air Force Base, TX 78235-5249

December 1997

NOTICES

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Office of Public Affairs has reviewed this technical report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

Stefan H. Constable

STEFAN H. CONSTABLE, Ph.D.
Project Scientist

JOHN G. MEYER, Lt Col, USAF, MC, FS
Chief, Office for Prevention & Health Services

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1997	Final (June-August 1996)

4. TITLE AND SUBTITLE

Modifications to the Air Force Cycle Ergometry Protocol: Impact on Pass, Fail, and Invalid Outcomes

5. FUNDING NUMBERS

6. AUTHOR(S)

Pete Flatten, Gerald De Wolfe, Stefan Constable, Ryan O'dowd, Dean Richardson, Susan Chao, Melissa Hite

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Armstrong Laboratory (AFMC)
Office for Prevention and Health Services Assessment (OPHSA)
2602 Doolittle Road
Brooks AFB TX 78235-5249

8. PERFORMING ORGANIZATION

AL/PS-TR-1997-0165

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING

11. SUPPLEMENTARY NOTES

This is the second report on results of a study of invalid outcomes from the Air Force's cycle ergometry fitness testing program.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Two modifications of the current Air Force cycle ergometry assessment protocol (ProtA) were compared to determine if invalid outcomes could be decreased. The first (ProtB) added a minute to the three workload (WL) progression stages, allowing more time to achieve a steady state. The second (ProtC) decreased the heart rate (HR) range for an increase in WL by 10 beats to make it more rigorous to increase the WL (a major cause of invalid outcomes). Ninety-three service members completed all three assessments, which were performed in a random order. ProtA had a pass rate of 60%, a fail rate of 8%, and an invalid rate of 32%. ProtB significantly reduced invalid outcomes and, therefore, increased the pass rate (pass = 73%, fail = 7%, and invalid = 20%). ProtC increased the pass rate to 66%, decreased the invalid rate to 26%, but was not significantly different than either ProtA or ProtB. VO2max scores between the protocols were not different. Decreased invalid outcomes in ProtB appear to result from a lower WL, which reduced the HR. The modified protocols reduced, but did not eliminate, invalid outcomes. Further research into a more automated system and/or more robust submaximal CE protocols is encouraged.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Cycle ergometry, US Air Force, invalid test results, fitness assessment, assessment protocol	25
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89) Prescribed by ANSI Std Z-39-18
298-102 COMPUTER GENERATED

Contents

Figures

Tables

SUMMARY

The purpose of this study was to compare the Pass, Fail, and Invalid rates of the current Air Force (AF) cycle ergometry (CE) assessment, designated Protocol A (ProtA), with two distinctly adjusted protocols, designated Protocol B (ProtB) and Protocol C (ProtC). ProtB lengthened by one minute each of the three stages at which a workload (WL) progression may occur, thereby allowing more time to achieve a steady state heart rate (HR). ProtC altered the computer logic to make it more difficult (i.e., required a lower HR response) for a subject to receive a 0.5 or 1.0 kilopond (Kp) increase in WL. Primarily, AF service members with a Category 1 Invalid outcome (HR greater than 85% of predicted maximum) on their last fitness test volunteered to complete the three assessments. Subjects were randomly assigned to a testing order (Protocols A, B, or C). One hundred and two subjects volunteered; 93 completed all three assessments. Major findings are outlined below.

Table 1: Frequency distribution of outcomes (Pass/Fail/Invalid) by protocol and predicted $\dot{V}O_2$max scores

Protocol	Pass	Fail	Invalid	Total	Predicted $\dot{V}O_2$max ($ml \cdot kg^{-1} \cdot min^{-1}$)*
A	56 (60%)	7 (8%)	30 (32%)	93 (100%)	36.8 ± 1.1
B	68 (73%)**	6 (7%)	19 (20%)**	93 (100%)	37.3 ± 1.1
C	61 (66%)	8 (8%)	24 (26%)	93 (100%)	37.3 ± 1.1

*Values = Mean ± Standard Error
**Significantly different from A ($p<.05$)
$\dot{V}O_2$max = maximal oxygen consumption

- ProtB had a significantly lower Invalid rate than ProtA ($p<.05$). There were no differences between ProtA and ProtC, or between ProtB and ProtC.
- Fifty-two percent of the subjects received a valid score on all three assessments. Twenty-nine percent received one Invalid, 8% received two Invalids, and 11% were assessed an Invalid outcome on all three protocols.
- For ProtA, 27 (90%) of the Invalids were attributed to Category 1, two (7%) to Category 2, and one (3%) was due to variable HR. In ProtB, only 13 (68%) of the Invalids were due to Category 1, one (5%) was due to Category 2, and five (27%) were due to Category 3. Finally, in ProtC 23 (96%) of Invalids were due to Category 1, there were zero Category 2 Invalids, and only one (4%) was a Category 3.
- There were no differences between protocols for the mean predicted $\dot{V}O_2$max.

Comparing the WL change in ProtA to ProtB and ProtC, ProtB assessments ended with more subjects pedaling against a lower WL than did ProtC (31% for ProtB and 13% for ProtC, versus ProtA).

The data suggested that ProtB was more conducive to keeping the WL lower, thereby keeping HR lower and allowing those individuals who were at risk of receiving a Category 1 Invalid to receive a score. In relation to ProtA, ProtB had a

decreased WL 31% of the time and an increased WL 4% of the time. ProtC had a decreased WL only 13% of the time, and resulted in an increased WL in 17% of subjects. Approximately 37% of subjects who had received an Invalid outcome in ProtA received a valid score in ProtB at a lower WL. For the same comparison, only 13% were at a lower WL in ProtC.

To determine if changes in exercise habits influenced assessment outcomes, an exercise questionnaire was completed by 87 of the 93 subjects.

- Of 87 responses, only 13% (11 subjects) reported no aerobic activity. Fifty-three percent reported aerobic activity more than 3 to 5 times per week. Fifty-eight percent reported no change in activity patterns since their last assessment.
- Of those reporting a change in activity (37 of 87), 22% had begun aerobic exercise since their last CE assessment, and 46% had increased the intensity, duration, or frequency of their exercise.
- Thirteen percent reported having stopped aerobic exercise, and 19% had decreased exercise intensity, duration, or frequency. Twenty-two percent reported a weight gain (13%) or loss (9%) of more than 5 pounds between their last AF CE assessment and their first study CE assessment.
- No associations between assessment outcomes and questionnaire responses were significant.

In conclusion, when compared to ProtA, ProtB appeared to be significantly better than ProtC at allowing a Pass or Fail score to be assessed (i.e., decreased the Invalid rate). This was likely accomplished by giving the HR longer to plateau in the WL progression phase, which resulted in a lower final WL. The current findings did not appear to be influenced by changes in reported activity patterns. Moreover, individual day-to-day variance in HR and/or tester competence may have influenced the likelihood of being assessed a score, as suggested by the high percentage of subjects who received a score on ProtA, when an Invalid outcome was expected. Finally, modest protocol changes may have a limited impact on solving the Invalid outcome problem. A system that could automatically enter HR and change WL, independent of cadence, would likely go a long way toward minimizing the Invalid rate. Further research should also focus on more robust sub-maximal protocols.

Note: Two coding errors in the study data were found after the analyses and report were complete. Both were related to Protocol B. The test outcomes of two subjects for Protocol B were incorrectly coded as "Invalid." Both should have been coded as "Pass." Changing these two test outcomes from "Invalid" to "Pass" would only strengthen the conclusion that Protocol B had a significantly lower invalid rate than Protocol A. Since rectifying the data does not alter the conclusion, and an enormous effort would be required to re-analyze the data, no correction was made to the results as presented in this report.

INTRODUCTION

The need to accurately measure the fitness level of each Air Force (AF) member has been addressed with a submaximal cycle ergometry (CE) assessment.[1] For a submaximal assessment to predict maximal oxygen consumption ($\dot{V}O_2$max), there often must be a period during which the heart rate (HR) is assessed at steady state. Currently, the HR must not vary more than three beats per minute (bpm) for the final two minutes of the AF assessment (i.e., minute 5 = 140 bpm and minute 6 = 143 bpm), after a minimum of 6 minutes at the same WL. If it does, an additional minute is added, which is then compared to the previous two minutes. $\dot{V}O_2$max is basically predicted from the steady state HR elicited by a set WL. During the assessment, the HR can range from a minimum of 125 bpm to a maximum of 85% of HR maximum (HRm; 85% of HRm is 220 minus age times .85). If an individual's HR response falls outside this range, $\dot{V}O_2$max may not be predicted as accurately.

There are three possible outcomes of the AF fitness assessment: Pass, Fail, or Invalid. When an individual's HR response falls outside of the designated range, or is not in steady state, the assessment is classified as an Invalid (Category 1, 2, or 3; originally there were seven Invalid categories). The CE assessment now too often results in an Invalid outcome, and no "score" is assessed. The subject must then be re-assessed, and a score is too frequently not derived on the re-test.

Anecdotal evidence from fitness assessment personnel suggested a majority of Invalid assessments were due to HR exceeding 85% of predicted maximum (Category 1). Excessive Invalid assessments, and the resulting need for re-assessment, are an unwanted drain on manpower and resources, as well as morale. A previous study by the AF Fitness Program[2] suggested that subjects who received a Category 1 Invalid outcome may have the greatest potential to receive a score, after an adjustment to the current protocol. Therefore, the purpose of this study was to compare the Pass, Fail, and Invalid rates of the current AF CE assessment (Protocol A [ProtA]) to those for two modified protocols (Protocol B and Protocol C; see Appendix A). Protocol B (ProtB) lengthened each of the three stages at which workload (WL) progression occurs by one minute, thereby allowing more time to achieve a steady state HR. Protocol C (ProtC) altered the WL response to the subject's HR, making it more rigorous for a subject to receive a 0.5 or 1.0 kilopond (Kp) WL progression (i.e., lowering the minimum HR needed to receive a WL increase; see Appendix B).

METHODS

Archived information on AF submaximal CE assessments from three San Antonio, Texas, bases (Brooks AFB, Kelly AFB, and Randolph AFB) performed in early fiscal year 1996 was collected and analyzed.[2] AF service members with a Category 1 Invalid (HR exceeded 85% of predicted maximum) were identified as potential subjects. These service members were recruited because of their likelihood to receive an Invalid on the current AF CE assessment.

For this study of the modified protocols, there were, again, three possible outcomes: Pass, Fail, or Invalid. Generally speaking, an Invalid outcome is assessed for one of three reasons: 1) the heart rate (HR) response falls outside of parameters set for the assessment (i.e., HR is not in a steady state, is above 85% of maximum, or falls below 125 bpm), 2) the subject requests termination, or 3) an error occurs due to equipment failure or administrator error. Originally, there were seven possible Invalid Category outcomes (Category 5 was deleted in April 1995):

1. HR exceeds 85% of HR maximum
2. HR does not reach 125 bpm in the last minute of the assessment
3. HR varies more than 3 bpm in the final 2 minutes
4. Subject cannot maintain 50 revolutions per minute (rpm)
5. Rating of Perceived Exertion (RPE) exceeds 15
6. Subject requests termination
7. Other

The identified service members were called, and the experiment was briefly described. If they volunteered to participate, they were scheduled for testing. One hundred and two subjects volunteered; 93 completed all three assessments. All subjects selected the time and day of testing from a schedule of available slots. Their CE protocol order was randomly assigned, based on the order in which they were scheduled (Table 1 in "Summary," page 1).

Upon arriving at the test site, which was their normal testing location (the base Health and Wellness Center), each subject read and signed a consent form, and filled out an exercise history questionnaire. Subjects were prepared for all three CE assessments using the standard preparation protocol found in the USAF Manual for Unit Fitness Program Manager.[3] Height and weight were measured while the subjects were either barefoot or wearing socks. Height was determined by using a measuring tape and a straight edge placed against the subject's head and wall. Weight was recorded with a standard hospital scale. Two pounds were deducted from the weight for a subject in workout clothes; three pounds were subtracted for a subject in battle dress uniform. Height was measured only on the first visit, while weight was measured before each assessment. A Polar HR monitor was placed on the subject. All assessments were run in FitSoft.[4] The only changes were those in the computer

software logic (Appendix A) that distinguished each protocol. All assessments were conducted by two master's-level exercise physiology graduate students who were thoroughly familiar with all assessment procedures.

ProtA (the current AF protocol) has two minutes of warm-up (25 Watts), followed by three minutes during which the WL may increase, depending on the HR. Six minutes of steady-state exercise at the same WL are accrued, and the final two HR are compared. If HR are within a three-beat range, the assessment is stopped and a $\dot{V}O_2$max score is assessed. If the last two HR are not within the three-beat range, an additional minute of exercise is completed, and the final three HR are compared. If none of the final three minutes are within the three-beat range, a Category 3 Invalid is assessed.

The ProtB algorithm employed two changes: 1) the three WL progression stages were extended by one minute each (the HR, therefore, had an extra minute per WL period to adjust), and 2) no minute of exercise was added if the final two HR (after 6 minutes of steady state) exceeded the three-beat range. Since the WL progression stages had up to three extra minutes, the additional minute was deemed unnecessary. Alternatively, if the HR varied by more than 3 bpm between the fifth and sixth minutes of steady state exercise, the HR for the fourth minute was manually compared. If any of the three were within the three-beat range, a score was calculated.

ProtC followed the same algorithm as ProtA, with one change: a criteria lowered by 10 bpm was used in minutes three and four of the WL progression to assess WL increases (i.e., a lower minimum HR was needed to receive a WL increase).[2] For example, in the current AF protocol, a 33-year-old subject with a HR of 102 bpm at minute 3 would receive a WL progression of 1 Kp. In ProtC, the same individual would receive a WL progression of .5 Kp, thereby keeping the HR lower. (See Appendix B for HR criteria.)

Subjects performed all three protocols on a Monarch 818E cycle ergometer. Assessments were separated by a minimum of one day, though not necessarily 24 hours. Subjects were also advised to refrain from vigorous exercise and caffeinated foods and beverages in the 12 hours preceding testing.

Data were transferred to Microsoft Excel 5.0 from hard-copy data sheets. Analysis was performed using the SAS software system. Both numeric and categorical outcomes were measured. Numeric outcomes were compared across protocols using one-way analysis of variance (ANOVA), repeated-measures design (applying the GLM procedure with the "REPEATED" statement in SAS). The weighted least square (WLS) method implemented in the SAS CATMOD procedure was used to analyze categorical repeated measures. Independent numeric outcomes were analyzed using a one-way ANOVA between design groups. When ANOVA revealed significant results, Tukey's Honest Significant Difference (HSD) multiple comparison test provided in the GLM procedure was performed, to help determine which pairs of groups were significantly different. Depending on the sample size, either the Pearson chi-square test or the exact test was used for the general association of categorical data. Pearson product-moment correlation coefficients (symbolized as "r") assessed bivariate linear relationships. All significance testing was done at the 0.05 level.

RESULTS

SUBJECT CHARACTERISTICS

One hundred and two individuals voluntarily participated in this study. Nine did not complete the three required assessments, resulting in 93 complete records (76 males and 17 females) upon which the analysis was based. Male and female data were combined because of the low number of female volunteers. Ninety-one percent of the subjects (n=85) had an Invalid outcome in the 1996 fiscal year. Ages ranged from 21 to 50; the mean age for both males and females was 35. Subject characteristics are described in Table 2. The order of the three protocols in which the subjects were tested was randomly assigned. Table 3 lists the number and percentage of subjects placed in each protocol pattern.

OUTCOMES

Table 1 (see "Summary," page 1) shows the Pass, Fail, and Invalid rates for each protocol, while Table 4 breaks down the Invalid outcomes by category. For ProtA, 56 (60%) of the subjects passed, seven (8%) failed, and 30 (32%) received an Invalid. For ProtB, 68 (73%) of the subjects passed, six (7%) failed, and 19 (20%) received an Invalid. And for ProtC, 61 (66%) passed, eight (8%) failed, and 24 (26%) received an Invalid. Protocol B was significantly different from A ($p<.05$). There were no differences between ProtA and ProtC, or between ProtB and ProtC.

The patterns of outcomes by protocol order (A, B, and C) are displayed in Table 5. Fifty-two percent of the sample received scores from all three assessments; the remainder received an Invalid on at least one of the protocols. The Invalid rates for protocols A, B, and C were 32%, 20%, and 26%, respectively. Pairwise comparisons of the Invalid rates between the three protocols show that the Invalid rate for ProtB was significantly lower than the rate for ProtA ($p<.05$). The Pass rates for protocols A, B, and C were 60%, 73%, and 66%, respectively (Table 1). A significantly higher Pass rate for ProtB was found, when compared to ProtA. Similar Fail rates of about 8% were observed for all three protocols (Table 1).

Twenty-nine percent of subjects received one Invalid, 8% received two Invalids, and 11% received an Invalid outcome on all three protocols (Table 5). For ProtA, 27 (90%) of Invalids were attributed to Category 1, two (7%) to Category 2, and one (3%) to Category 3 (Table 4). In ProtB, only 13 (68%) Invalids were due to Category 1, one (5%) was due to Category 2, and five (27%) were due to Category 3. In ProtC, 23 (96%) Invalids were due to Category 1, there were no Category 2 Invalids, and only one (4%) was due to Category 3.

ENDING WORKLOAD

The frequency distribution of WL change from ProtA to ProtB by outcome is presented in Table 6A. The frequency distribution of outcome by WL change from ProtA to ProtB is displayed in Table 6B. Overall, 60 of 93 subjects (65%) received the same WL in ProtA and ProtB, 31% of subjects received a reduced WL for ProtB, and 4%, received an increased WL for ProtB. A significant association between change of WL and outcome was revealed ($p<0.05$). Sixty-three percent (19 of 29) of subjects who received Invalid assessment outcomes from ProtA received valid outcomes from ProtB. Subjects who received Invalid outcomes from ProtA had a significantly higher ($p<0.05$) likelihood of receiving a valid outcome when they received a reduced WL (92%, or 11 of 12), rather than if the WL remained the same (44%, or 8 of 18). Among those who received a valid outcome from ProtA, 100% received a valid outcome from ProtB when WL was reduced, 83% received a valid outcome when WL stayed the same, and 75% received a valid outcome when WL increased.

Tables 6A and 7 demonstrate that, in relation to ProtA (of those assessments ending in a score; row one of Tables 6A and 7), ProtB decreased the WL 31% of the time and increased WL 4% of the time, while ProtC decreased WL only 13% of the time and increased WL 15% of the time. Of those who received an Invalid in ProtA and a score in ProtB (n=19), 11 were at a lower WL in ProtB, while eight remained at the same WL. In contrast, of those who received an Invalid in ProtA and a score in ProtC (n=15), four were at a lower WL on ProtC, 10 were at the same WL, and one was at a greater WL. Invalid outcomes described in Table 4 support these findings.

QUESTIONNAIRE

Exercise questionnaire responses are summarized in Table 8 (see Appendix C for questions). Of 87 completed questionnaires, only 13% (11 subjects) reported no aerobic activity. Fifty-three percent reported aerobic activity more than 3 to 5 times per week. Fifty-seven percent reported no change in activity patterns since their last assessment. Of those reporting a change in activity (37 of 87), 22% indicated that they had begun aerobic exercise since their last CE assessment, and 46% had increased the intensity, duration, or frequency of their exercise. Conversely, 13% had stopped exercising aerobically, and 19% had decreased its intensity, duration, or frequency. Thirteen percent of respondents reported a weight gain of more than five pounds, while 9% lost more than five pounds, between their last official AF assessment and the first day of the study.

$\dot{V}O_2$MAX

The predicted $\dot{V}O_2$max scores for each protocol were not significantly different (Table 1). Variations in $\dot{V}O_2$max scores between protocols by the change in ending WL are noted in Table 9. The majority of subjects completed their assessments at the same WL. Although not significant, those whose WL ended at a higher resistance, and therefore a higher HR, tended to score about a 2.0 $ml \cdot kg^{-1} \cdot min^{-1}$ higher $\dot{V}O_2$max (Table 9).

DISCUSSION

Eighty-five of the 93 volunteers had received an Invalid outcome in the previous testing year. They were selected so that the effect of a different protocol could be measured on those it was supposed to impact: individuals predisposed to receiving an Invalid outcome and, in particular, those who had received multiple Invalids. A surprising finding was that 60% of all volunteers passed, 8% failed, and 32% had an Invalid outcome in ProtA. These percentages are near those expected for the AF population as a whole, as found in a recent review of Pass, Fail and Invalid rates from five AF bases, a sample assumed to be representative of the AF (9437 assessments).[2] There was a 74.0% Pass rate, a 9.6% Fail rate, and a 16.4% Invalid rate in that sample. What is interesting in comparing Pass, Fail and Invalid percentages between this study's three protocols is the low Invalid rate for all protocols, even though the subject population was predicted to be predisposed to receiving an Invalid. This could be due to differences in anxiety, stress, individual variation (i.e., day-to-day HR variation), response to a training regimen, and/or tester proficiency. Others have acknowledged that factors other than chronic physical training have potential influence on submaximal cycle ergometry testing.[5]

Analysis of the Pass, Fail, and Invalid rates of the 93 subjects showed that ProtB resulted in significantly fewer Invalid outcomes, and therefore more valid scores, than the two other protocols (Tables 1 and 4). ProtC also lowered the Invalid rate, but only ProtB was significantly different ($p<.05$) than ProtA. Of those who had one Invalid outcome from the three assessments, 14 occurred in ProtA, 6 occurred in ProtB, and 8 occurred in ProtC (see bold figures in Table 5). There were no significant differences in the predicted $\dot{V}O_2$max scores between the protocols (Table 1), suggesting that modifications made to the protocol did not change the reliability of regression equations used to predict aerobic capacity.

The data suggest that ProtB was more conducive to keeping the WL lower, thereby keeping HR lower and allowing a score to be assessed (reducing the incidence of Category 1 Invalids). As found by DeWolfe et al.,[2] the greatest number of Invalids is due to members exceeding 85% of their predicted maximal HR (Category 1 Invalid). By keeping the WL lower, ending HR was reduced, which kept more subjects' HR below 85% of predicted HR max.

To determine if changes in exercise, activity patterns, and/or weight had an impact on results, a questionnaire on subjects' exercise habits since their last assessment and changes in body composition was completed (Table 8 and Appendix C). No association between assessment outcomes and survey question responses was significant. In general, questionnaires are subjective and difficult to interpret. It appeared to the investigators that service members were concerned about the use of questionnaire responses in tracking and compliance. Therefore, the questions may have been answered, to some extent, in a way that showed the service member to be more active or more fit, and may not have reflected true exercise habits. If the responses indeed were accurate, we would have expected those who reported regular exercise to have had a higher pass rate, and vice versa. Since this did not occur, the questionnaire's validity could be debated.

Observations by the investigators suggested that those who were anxious or under stress tended to be more likely to receive an Invalid outcome. Anxiety and stress, by themselves, cause HR to rise and, therefore, might be suggested to contribute to the Category 1 Invalid rate. Conversely, elevated resting HR brought on by stress is assumed generally to be overridden by demands placed on the heart during exercise, so anxiety will not significantly contribute to Category 1 Invalids.

To examine whether anxiety played a role in the Invalid rate, first-day (first-assessment) results were compared to those from the second and third assessments. (It was assumed that the first assessment would be the one most likely to induce anxiety.) Frequency distribution of outcomes by assessment day (first, second, or third) and by protocol is presented in Table 10. Overall (see the "Total" row, Table 10), the Invalid rate (31%) of the first assessment was the highest. A significant difference ($p<0.05$) in the Invalid rate between the first and second assessments (31% vs. 20%) was found. The third assessment outcome was similar to the first. When the rates are examined by protocol, the distribution of Invalid rates by day for ProtA was different from that for ProtB and ProtC. The Invalid rates for the first (39%) and third assessments (42%) for ProtA were similar; both were significantly higher ($p<0.05$) than the Invalid rate for the second assessment (16%). We consider this to be an anomaly. Day-to-day variability in stress level, hydration status, etc., probably accounted for the differences. Only ProtC displayed a pattern that supports the hypothesis that anxiety influences the assessment outcome.

It is possible that, during submaximal CE testing, the stress of exercise is not always sufficient to "cover" the stress of assessment, especially if it is completed at a low WL relative to $\dot{V}O_2max$. Reducing subject anxiety before testing may decrease the Invalid rate overall, but the potential impact is thought to be negligible. Summary statistics of starting and ending HR by protocol and WL are displayed in Table 11. Significant ($p<.0001$) negative correlations between starting HR and ending WL were found in all three protocols. The correlation coefficients between starting HR and ending WL were: -0.55 for ProtA, -0.49 for ProtB, and -0.68 for ProtC (see Figures 1, 2, and 3). In other words, a higher starting HR was more likely to result in a lower ending WL, and vice versa. However, as indicated in Table 12, there was no difference in starting HR

between those who received a score and those who did not. This further supports the statistics showing that starting HR does not predispose one to an Invalid result. These data suggest that a higher starting HR would result in a lower WL, which generally would facilitate achieving a valid outcome.

CONCLUSIONS

ProtB was significantly better at reducing the Invalid rate (i.e., increasing the Pass/Fail rate) than ProtC. It appears to achieve this by giving the HR longer to plateau in each WL progression phase, often resulting in a lower final WL. While starting HR was negatively correlated with ending WL, the relationship between starting HR and assessment outcome was not significant (Table 12). Anxiety appears to play a minimal role in eliciting Invalid outcomes. The high percentage of subjects receiving a score on ProtA, despite expectations from past assessments that they would receive an Invalid, suggests that factors other than just WL decreases contribute to the high Invalid rate experienced in the field. For example, test administrators' competence and/or experience cannot be ruled out as significantly influencing assessment outcomes. This would suggest that more rigorous FAM training and supervision could reduce the likelihood of Invalid outcomes, beyond that which can be achieved by protocol changes alone. The investigators' position is that individual, day-to-day variance in HR, physiological status, and tester competence all influence the probability of being assessed a CE score. Stress and anxiety do not appear to contribute significantly to the Invalid rate.

REFERENCES

1. Pollock, ML, et al. 1994. *The cross-validation of the United States Air Force submaximal cycle ergometer test to estimate aerobic capacity* (AL/CF-TR-1994-0046). Brooks AFB, Texas: Crew Technology Division, Armstrong Laboratory, Crew Systems Directorate.

2. DeWolfe, G, et al. 1997. *Invalid cycle ergometry assessment outcomes at five Air Force bases* (AL/PS-TR-1997-0056). Brooks AFB, Texas: Clinical Preventive Services Division, Armstrong Laboratory, Preventive Services Directorate.

3. Air Force Fitness Program. 1995. *Air Force fitness program manual for installation fitness program manager.* Brooks AFB, Texas: Air Force Fitness Program Office.

4. Air Force Fitness Program. 1996. *Performance specification: Air Force fitness program (AFFP) fitness assessment software (FPAS)* (YAE-9501-01). Brooks AFB, Texas: Engineering Division, Human Systems Program Office.

5. Howley, E. 1997. You asked for it: question authority. *ACSM Health & Fitness Journal* 1(5):13,45

FIGURES

(Note: n = valid outcomes only)

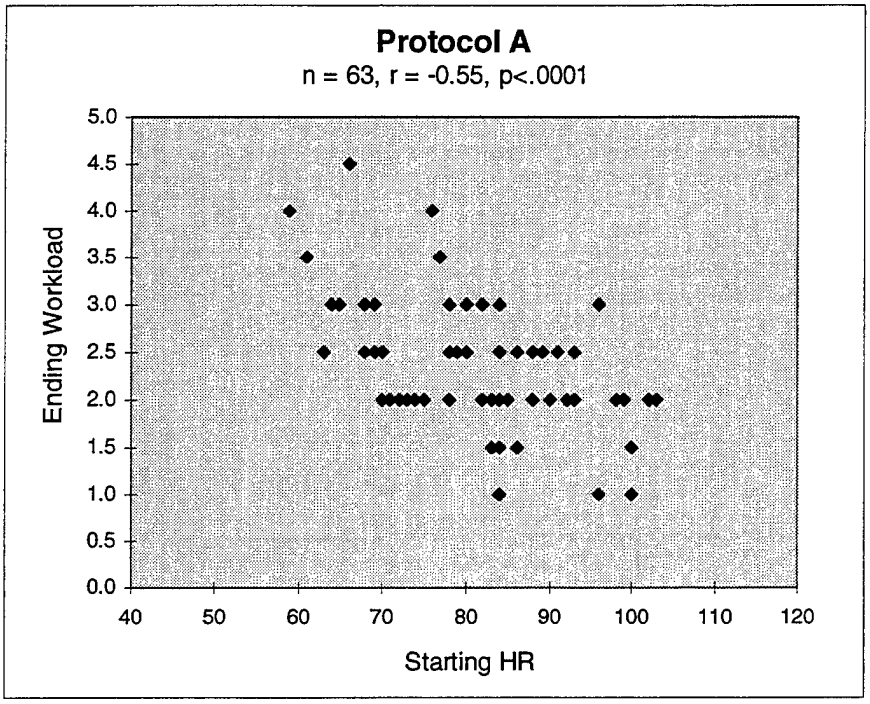

Figure 1: Protocol A, starting heart rate vs. ending workload

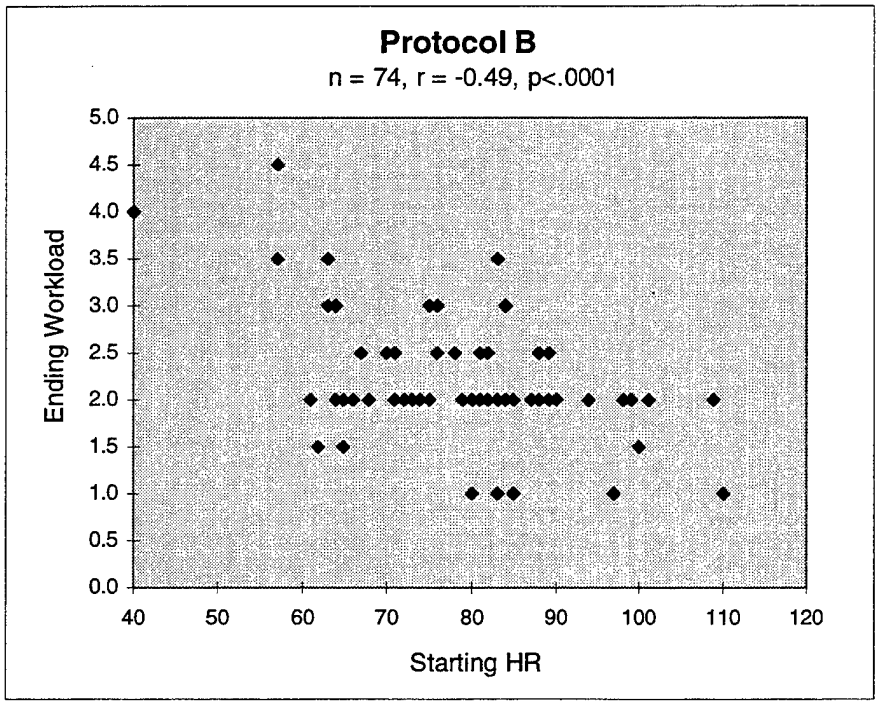

Figure 2: Protocol B, starting heart rate vs. ending workload

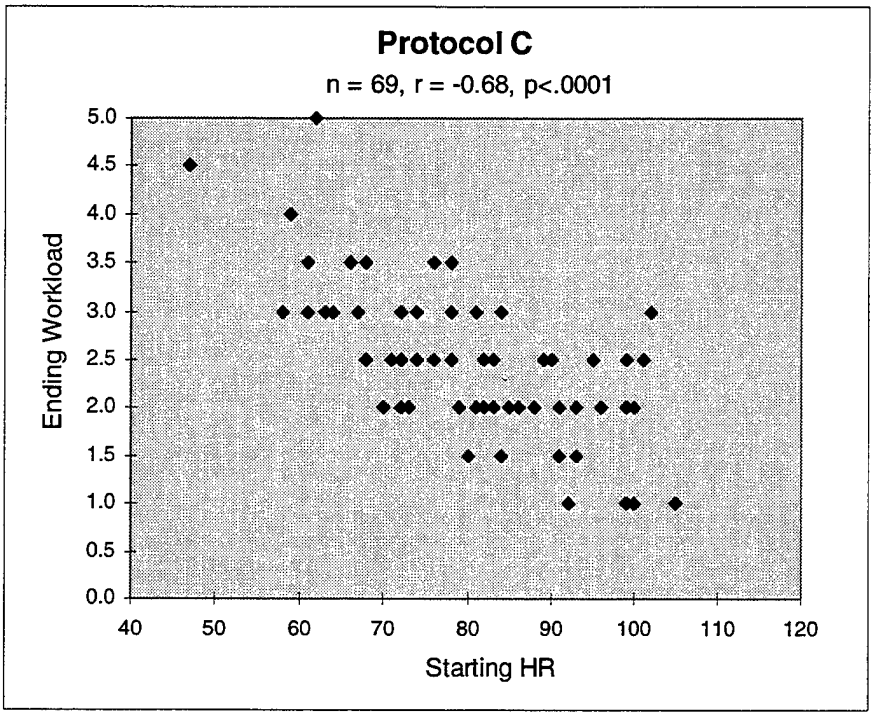

Figure 3: Protocol C, starting heart rate vs. ending workload

TABLES

Table 2: Descriptive statistics of age, height and weight by gender

Variable	Gender	n	Mean ± Standard Error	Minimum	Maximum
Age (year)	Male	76	35 ± 0.8	21	50
	Female	17	35 ± 1.5	26	44
Weight (lb.)	Male	76	175 ± 2.8	127	244
	Female	17	137 ± 5.9	89	174
Height (in.)	Male	76	70 ± 0.3	63	74
	Female	17	64 ± 0.6	59	68

Table 3: Frequency distribution of the protocol patterns

Pattern	Frequency (percent)
ABC	17 (18%)
ACB	14 (15%)
BAC	15 (16%)
BCA	16 (17%)
CAB	16 (17%)
CBA	15 (16%)
Total	93 (100%)

Table 4: Frequency distribution of the outcomes (Pass/Fail/Invalid) by protocol

Protocol	Pass	Fail	Invalid	Type of Invalid			Total
A	56 (60%)	7 (8%)	30 (32%)				93 (100%)
				HR too high	27	(90%)	
				HR too low	2	(7%)	
				HR unstable	1	(3%)	
B	68 (73%)*	6 (7%)	19 (20%)**				93 (100%)
				HR too high	13	(68%)	
				HR too low	1	(5%)	
				HR unstable	5	(27%)	
C	61 (66%)	8 (8%)	24 (26%)				93 (100%)
				HR too high	23	(96%)	
				HR too low	0	(0%)	
				HR unstable	1	(4%)	

* Significantly different from the pass rate of ProtA

** Significantly different from the invalid rate of ProtA

Table 5: Frequency distribution of the pattern of the outcomes

Outcome (ABC)	Frequency	Percent
000*	48	52%
001	**7**	**8%**
010	**6**	**6%**
011	2	2%
100	**14**	**15%**
101	5	5%
110	1	1%
111	10	11%
Total	93	100%

*0=valid (P or F), 1=invalid

Table 6A: Frequency distribution of workload change from Protocol A to Protocol B by outcome

	Change of Workload							
Outcome	Decreased		No Change		Increased		Total	
Valid A, Valid B	17	(31%)	35	(64%)	3	(5%)	55	(100%)
Valid A, Invalid B	0	(0%)	7	(88%)	1	(12%)	8	(100%)
Invalid A, Valid B	11	(58%)	8	(42%)	0	(0%)	19	(100%)
Invalid A, Invalid B	1	(9%)	10	(91%)	0	(0%)	11	(100%)
Total	29	(31%)	60	(65%)	4	(4%)	93	(100%)

Table 6B: Frequency distribution of outcome by workload change from Protocol A to Protocol B

	Change of Workload					
Outcome	Decreased		No Change		Increased	
Valid A, Valid B	17	(100%)	35	(83%)	3	(75%)
Valid A, Invalid B	0	(0%)	7	(17%)	1	(25%)
	17	(100%)	42	(100%)	4	(100%)
Invalid A, Valid B	11	(92%)*	8	(44%)	0	(0%)
Invalid A, Invalid B	1	(8%)	10	(56%)	0	(0%)
	12	(100%)	18	(100%)	0	(100%)

* Significantly different from the valid rate of those whose workload did not change

Table 7: Frequency distribution by outcome of workload change from Protocol A to Protocol C

	Change of Workload							
Outcome	Decreased		No Change		Increased		Total	
Valid A, Valid C	7	(13%)	39	(72%)	8	(15%)	54	(100%)
Valid A, Invalid C	0	(0%)	4	(44%)	5	(56%)	9	(100%)
Invalid A, Valid C	4	(27%)	10	(67%)	1	(7%)	15	(100%)
Invalid A, Invalid C	1	(7%)	12	(80%)	2	(13%)	15	(100%)
Total	12	(13%)	65	(70%)	16	(17%)	93	(100%)

Table 8: Descriptive statistics of responses to the initial subject exercise questionnaire

Question	n	n Missing	Category	Frequency	Percent
AFIP/MFIP	87	6	Yes	26	30%
			No	61	70%
Aerobic activities (times/week)	87	6	No activity	11	13%
			1-2	30	35%
			3-5	39	45%
			6-7	7	8%
Changed activity pattern	87	6	Yes	37	43%
			No	50	57%
Activity pattern changed	37	56	Exercising aerobically		
			Began	8	22%
			Stopped	5	13%
			Increased	17	46%
			Decreased	7	19%
Status of weight	87	6	Not changed	68	78%
			Gained more than 5 lb.	11	13%
			Lost more than 5 lb.	8	9%
Sleeping pattern changed	87	6	No	79	91%
			Yes/less sleep	6	7%
			Yes/more sleep	2	2%
Status of tobacco use	87	6	Do not use	61	70%
			Stopped	4	5%
			Started	2	2%
			Not changed	20	23%

Table 9: Summary statistics of change of $\dot{V}O_2$max and change of ending HR by workload change, for individuals with valid outcomes

Protocol	Workload	n	Change of $\dot{V}O_2$ (Mean ± Standard Error)	Change of ending HR (Mean ± Standard Error)
A to B	Decreased	17	-0.5 ± 0.8	-13.1 ± 1.7
	No Change	35	0.8 ± 0.4	-1.9 ± 1.0[§]
	Increased	3	2.7 ± 2.3	5.7 ± 6.3[§†]
A to C	Decreased	7	-1.7 ± 1.3	-7.7 ± 2.9
	No Change	39	0.5 ± 0.3	-1.3 ± 0.8[§]
	Increased	8	2.0 ± 1.2*	9.8 ± 1.8[§†]

* Significantly different from the mean change of $\dot{V}O_2$ of those who had decreased workload
[§] Significantly different from the mean change of ending HR of those who had decreased workload
[†] Significantly different from the mean change of ending HR of those whose workload did not change

Table 10: Frequency distribution of assessment outcomes by day and protocol

	1st Assessment				2nd Assessment				3rd Assessment				
Protocol	Valid		Invalid		Valid		Invalid		Valid		Invalid		Total
A	19	(61%)	12	(39%)*	26	(84%)	5	(16%)	18	(58%)	13	(42%)*	93
B	24	(77%)	7	(23%)	26	(81%)	6	(19%)	24	(80%)	6	(20%)	93
C	21	(68%)	10	(32%)	22	(73%)	8	(27%)	26	(81%)	6	(19%)	93
Total	64	(69%)	29	(31%)*	74	(80%)	19	(20%)	68	(73%)	25	(27%)	

* Significantly different from the invalid rate of the 2nd Assessment

Table 11: Summary statistics of starting and ending HR by protocol and ending WL (valid outcomes only)

Protocol	Ending Workload	n	Starting HR Mean ± Standard Error	Ending HR Mean ± Standard Error
A	1.0	4	95.0 ± 3.8	144.0 ± 5.0
	1.5	4	88.3 ± 4.0	144.5 ± 5.2
	2.0	23	83.7 ± 2.1	144.5 ± 1.8
	2.5	18	79.6 ± 2.0	145.1 ± 1.9
	3.0	9	76.2 ± 3.5	147.4 ± 3.0
	3.5	2	69.0 ± 8.0	155.0 ± 3.0
	4.0	2	67.5 ± 8.5	137.5 ± 4.5
	4.5	1	66.0 ± —	139.0 ± —
Correlation coefficient between ending WL and starting HR = -0.55 (p=0.0001) Correlation coefficient between ending WL and ending HR = 0.030 (p=0.82)				
B	1.0	6	88.7 ± 3.9	135.5 ± 2.8
	1.5	3	77.3 ± 11.9	143.7 ± 4.1
	2.0	37	83.6 ± 1.9	139.6 ± 1.2
	2.5	16	80.0 ± 2.7	142.8 ± 2.6
	3.0	7	75.1 ± 3.6	136.1 ± 3.2
	3.5	3	68.0 ± 4.2	140.7 ± 1.2
	4.0	1	59.0 ± —	132.0 ± —
	4.5	1	66.0 ± —	137.0 ± —
Correlation coefficient between ending WL and starting HR = -0.49 (p=0.0001) Correlation coefficient between ending WL and ending HR = 0.002 (p=0.98)				
C	1.0	5	99.2 ± 2.1	147.6 ± 5.1
	1.5	4	87.0 ± 3.0	145.0 ± 4.1
	2.0	25	85.0 ± 1.8	143.5 ± 1.8
	2.5	16	82.2 ± 2.5	149.7 ± 1.8
	3.0	11	73.1 ± 3.9	145.2 ± 2.6
	3.5	5	69.8 ± 3.2	143.8 ± 4.7
	4.0	1	59.0 ± —	152.0 ± —
	4.5	1	47.0 ± —	137.0 ± —
	5.0	1	62.0 ± —	145.0 ± —
Correlation coefficient between ending WL and starting HR = -0.68 (p=0.0001) Correlation coefficient between WL and ending HR = -0.013 (p=0.91)				

Table 12: Comparison of starting heart rates between subjects with valid and invalid outcomes by protocol

Protocol	Outcome	n	Starting HR (Mean ± Standard Error)	p-Value
A	Valid	63	81.2 + 1.4	0.66
	Invalid	30	79.8 + 2.8	
B	Valid	74	79.1 + 1.4	0.56
	Invalid	19	81.0 + 3.4	
C	Valid	69	81.2 + 1.5	0.01
	Invalid	24	73.3 + 2.3	

APPENDIX A: PROTOCOLS

WL = Workload

Protocol A: The original protocol for AF cycle ergometry testing

Protocol B: Lengthens each of the three stages at which workload progression occurs by one minute, thereby allowing more time to achieve a steady state HR

Protocol C: Alters the computer logic to make it more difficult for a subject to receive a 1.0 kilopond (Kp) or 0.5 Kp workload progression (i.e., lower the minimum HR needed to receive a workload increase)

Time (min)	A	B	C
-	Initial	Initial	Initial
1	Warm-up	Warm-up	Warm-up
2	Warm-up	Warm-up	Warm-up
3	WL Progression	Steady-state	WL Progression
4	WL Progression	WL Progression	WL Progression
5	WL Progression	Steady-state	WL Progression
6	Steady-state	WL Progression	Steady-state
7	Steady-state	Steady-state	Steady-state
8	Steady-state	WL Progression	Steady-state
9	Steady-state	Steady-state	Steady-state
10	Steady-state	Steady-state	Steady-state
11	Steady-state	Steady-state	Steady-state
12	Optional	Steady-state	Optional
13	-	Steady-state	-
14	-	Steady-state	-

Note: Subjects must complete six minutes of steady-state workload. This steady-state phase begins as soon as there is no further workload progression. Steady-state heart rate is designated by a heart rate in the final minute within ± 3 beats of the previous minute. Test will be extended by one minute for individuals not in steady-state during the final minute (Protocols A and C).

APPENDIX B: INITIAL WORKLOAD SETTINGS/HEART RATE PARAMETERS

Table B1: Initial workload setting for females, in Kp.

	Weight									
	<54.88 kg (<121 lb.)		<63.95 kg (<141 lb.)		<73.02 kg (<161 lb.)		<82.09 kg (<181 lb.)		<226.76 kg (<500 lb.)	
Exercise History	Active	Inactive	Active	Inactive	Active	Inactive	Active	Inactive	Active	Inactive
Age										
17 - 35	1.0	1.0	1.5	1.0	1.5	1.0	2.0	1.5	2.0	2.0
36 - 50	1.0	1.0	1.5	1.0	1.5	1.0	2.0	1.5	2.0	1.5
51 - 62	1.0	1.0	1.0	1.0	1.0	1.0	1.5	1.0	1.5	1.0
63 - 70	1.0	0.5	1.0	0.5	1.0	1.0	1.5	1.0	1.5	1.0

Table B2: Initial workload settings for males, in Kp.

	Weight									
	<59.41 kg (<131 lb.)		<68.48 kg (<151 lb.)		<82.09 kg (<181 lb.)		<100.23 kg (<221 lb.)		<226.76 kg (<500 lb.)	
Exercise History	Active	Inactive	Active	Inactive	Active	Inactive	Active	Inactive	Active	Inactive
Age										
17 - 35	1.5	1.0	2.0	1.5	2.0	2.0	2.5	2.0	2.5	2.5
36 - 50	1.5	1.0	1.5	1.5	2.0	2.0	2.0	2.0	2.5	2.0
51 - 62	1.5	1.0	1.5	1.0	2.0	1.5	2.0	1.5	2.0	1.5
63 - 70	1.0	1.0	1.5	1.0	1.5	1.0	1.5	1.5	1.5	1.5

Table B3: Heart rate parameters for workload progression (Protocols A and B)

	Workload Progression											
	+1 Kp			+0.5 Kp			0.0 Kp			Terminate Assessment		
Minute	3	4	5	3	4	5	3	4	5	3	4	5
Age												
17 - 30	<110	<110	<115	110-119	110-119	115-128	120-173	120-173	129-173			
31 - 40	<105	<105	<110	105-114	105-114	110-126	115-161	115-161	127-161	Invalid if		
41 - 50	<100	<100	<105	100-109	100-109	105-122	110-152	110-152	123-152	>85% of max.		
51 - 60	<100	<100	<105	100-109	100-109	105-120	110-144	110-144	121-144	heart rate		
61 - 70	<90	<90	<95	90-104	90-104	95-105	105-135	105-135	106-135			

Progression workload cycle changes.*

*Note: Heart rates used to determine workload progression are taken at the end of the minute. For example, minute 3 of the assessment is performed at the initial workload; the heart rate at the end of minute 3 determines the workload progression for minute 4 using the "Minute 3" workload progression column.

Table B4: Heart rate parameters for workload progression (Protocol C)

	Workload Progression											
	+1 Kp			+0.5 Kp			0.0 Kp			Terminate Assessment		
Minute	3	4	5	3	4	5	3	4	5	3	4	5
Age												
17 - 30	**<100**	**<100**	<115	**100-119**	**100-119**	115-128	120-173	120-173	129-173			
31 - 40	**<95**	**<95**	<110	**95-114**	**95-114**	110-126	115-161	115-161	127-161	Invalid if		
41 - 50	**<90**	**<90**	<105	**90-109**	**90-109**	105-122	110-152	110-152	123-152	>85% of max.		
51 - 60	**<90**	**<90**	<105	**90-109**	**90-109**	105-120	110-144	110-144	121-144	heart rate		
61 - 70	**<80**	**<80**	<95	**80-104**	**80-104**	95-105	105-135	105-135	106-135			

Progression workload cycle changes.*

*Note: Heart rates used to determine workload progression are taken at the end of the minute. For example, minute 3 of the assessment is performed at the initial workload; the heart rate at the end of minute 3 determines the workload progression for minute 4 using the "Minute 3" workload progression column.

APPENDIX C: INITIAL SUBJECT SURVEY

1) Are you currently participating in either the self-directed, unsupervised fitness improvement program (SFIP) or the monitored fitness improvement program (MFIP)?
   A) Yes
   B) No

2) Please indicate the number of times per week you currently participate in aerobic activities (running, walking, cycling, rowing, swimming, Nordic skiing...) lasting 20 minutes.
   A) No activity
   B) 1-2
   C) 3-5
   D) 6-7

3) Have you changed your activity pattern since taking your last cycle ergometry test?
   A) Yes.. If yes please proceed to question 4.
   B) No.. If no please proceed to question 6.

4) Please indicate date of last test: _____

5) How has your activity pattern changed since your last cycle ergometry test?
   A) I began exercising aerobically since I had my last cycle ergometry test.
   B) I have stopped exercising aerobically since I had my last cycle ergometry test.
   C) I have increased either the intensity, duration, or frequency per week of my aerobic program.
   D) I have decreased either the intensity, duration, or frequency per week of my aerobic program.

6) Please indicate the status of your weight since your last cycle ergometry test.
   A) My weight has not changed.
   B) I have gained more than 5 pounds.
   C) I have lost more than 5 pounds.

7) Have your sleeping patterns (average hours of sleep per night) changed since your last cycle ergometry test?
   A) No
   B) Yes... I am getting less sleep per night.
   C) Yes... I am getting more sleep per night.

8) Please indicate the status of tobacco use since your last cycle ergometry test.
   A) I do not use any tobacco products.
   B) I have stopped using tobacco products since my last cycle ergometry test.
   C) I have started using tobacco products since my last cycle ergometry test.
   D) My tobacco use has not changed.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

ANOVA	analysis of variance
bpm	beats per minute
CE	cycle ergometry
FAM	Fitness Assessment Monitor
FitSoft	cycle ergometry software
HR	heart rate
HRm	heart rate 85% of maximum (220 minus age times .85)
Kp	kilopond
Protocol A	regular AF cycle ergometry protocol
Protocol B	lengthened each of the three stages at which workload progression occurs by one minute
Protocol C	altered the computer logic to make it more difficult (i.e., required a lower heart-rate response) for a subject to receive a 0.5 or 1.0 kilopond (Kp) increase in workload
RPE	Rating of Perceived Exertion
$\dot{V}O_2max$	maximal oxygen consumption
WL	workload
WLP	workload progression